

**In the Specification:**

Please delete the paragraph beginning at page 5, line 21 of the specification.

Please delete the paragraph beginning at page 5, line 23 of the specification.

Please delete the paragraph beginning at page 5, line 25 of the specification

Please amend the paragraph beginning at page 5, line 28 of the specification to read as follows:

--Figure 6 shows an exemplary descriptor.--

Please amend the paragraph beginning at page 5, line 29 of the specification to read as follows:

--Figure 7 shows a set of e-value scores for ferritin.--

Please delete the paragraph beginning at page 6, line 1 of the specification.

Please delete the paragraph beginning at page 6, line 3 of the specification.

Please amend the paragraph beginning at page 6, line 5 of the specification to read as follows:

--Figure 8 shows a representative lookup table used in Q-compare or CompareOverWins.--

Please amend the paragraph beginning at page 6, line 7 of the specification to read as follows:

--Figure 9 shows a representative block diagram of a program called RevComp.--

**DOCKET NO.: IBIS-0012**

Please amend the paragraph beginning at page 6, line 8 of the specification to read as follows:

--Figure 10 shows a representative flow chart showing preferred steps of a preferred database search strategy for ortholog finding.--

Please delete the paragraph beginning at page 6, line 10 of the specification.

Please delete the paragraph beginning at page 6, line 12 of the specification.

Please amend the paragraph beginning at page 6, line 14 of the specification to read as follows:

--Figure 11 shows a representative flow scheme showing preferred steps for a preferred SEALS strategy.--

Please delete the paragraph beginning at page 6, line 16 of the specification.

Please amend the paragraph beginning at page 6, line 18 of the specification to read as follows:

--Figure 12 represents a genetic map showing a conserved iron response element in the 5' UTR of mouse and human ferritin.--

Please delete the paragraph beginning at page 6, line 20 of the specification.

Please delete the paragraph beginning at page 6, line 22 of the specification.

Please delete the paragraph beginning at page 6, line 24 of the specification.

Please delete the paragraph beginning at page 6, line 25 of the specification.

Please amend the paragraph beginning at page 6, line 26 of the specification to read as follows:

--Figure 13 shows representative flow scheme showing preferred steps for a preferred Structure Predictor strategy.--

Please delete the paragraph beginning at page 6, line 28 of the specification.

Please delete the paragraph beginning at page 6, line 30 of the specification.

Please amend the paragraph beginning at page 7, line 1 of the specification to read as follows:

--Figure 14 shows a representative structure drawing of ferritin 5' UTR.--

Please delete the paragraph beginning at page 7, line 2 of the specification.

Please delete the paragraph beginning at page 7, line 3 of the specification.

Please delete the paragraph beginning at page 7, line 5 of the specification.

Please delete the paragraph beginning at page 7, line 6 of the specification.

Please delete the paragraph beginning at page 7, line 7 of the specification.

Please delete the paragraph beginning at page 7, line 9 of the specification.

Please delete the paragraph beginning at page 7, line 11 of the specification.

Please delete the paragraph beginning at page 7, line 12 of the specification.

Please delete the paragraph beginning at page 7, line 14 of the specification.

Please delete the paragraph beginning at page 7, line 15 of the specification.

Please delete the paragraph beginning at page 7, line 16 of the specification.

Please delete the paragraph beginning at page 7, line 18 of the specification.

Please delete the paragraph beginning at page 7, line 20 of the specification.

Please delete the paragraph beginning at page 7, line 22 of the specification.

Please delete the paragraph beginning at page 7, line 23 of the specification.

Please delete the paragraph beginning at page 7, line 25 of the specification.

Please delete the paragraph beginning at page 7, line 27 of the specification.

Please delete the paragraph beginning at page 7, line 28 of the specification.

Please delete the paragraph beginning at page 7, line 30 of the specification.

Please delete page 8.

Please delete the paragraph beginning at page 9, line 1 of the specification.

Please delete the paragraph beginning at page 9, line 3 of the specification.

Please delete the paragraph beginning at page 9, line 5 of the specification.

Please amend the paragraph beginning at page 9, line 7 of the specification to read as follows:

--Figure 15 shows a representative mass-spec structure probe analysis of region 1 of ornithine decarboxylase 3' UTR.--

Please delete the paragraph beginning at page 9, line 9 of the specification.

Please delete the paragraph beginning at page 9, line 11 of the specification.

Please delete the paragraph beginning at page 9, line 13 of the specification.

Please delete the paragraph beginning at page 9, line 15 of the specification.

Please delete the paragraph beginning at page 9, line 17 of the specification.

Please delete the paragraph beginning at page 9, line 19 of the specification.

Please delete the paragraph beginning at page 9, line 21 of the specification.

Please delete the paragraph beginning at page 9, line 23 of the specification.

Please delete the paragraph beginning at page 9, line 25 of the specification.

Please delete the paragraph beginning at page 9, line 27 of the specification.

Please delete the paragraph beginning at page 9, line 29 of the specification.

Please delete the paragraph beginning at page 9, line 30 of the specification.

Please delete the paragraph beginning at page 10, line 1 of the specification.

Please delete the paragraph beginning at page 10, line 3 of the specification.

Please delete the paragraph beginning at page 10, line 5 of the specification.

Please delete the paragraph beginning at page 10, line 6 of the specification.

Please delete the paragraph beginning at page 10, line 8 of the specification.

Please delete the paragraph beginning at page 10, line 10 of the specification.

Please delete the paragraph beginning at page 10, line 11 of the specification.

Please delete the paragraph beginning at page 10, line 12 of the specification.

Please delete the paragraph beginning at page 10, line 14 of the specification.

Please delete the paragraph beginning at page 10, line 16 of the specification.

Please delete the paragraph beginning at page 10, line 17 of the specification.

Please delete the paragraph beginning at page 10, line 18 of the specification.

Please delete the paragraph beginning at page 10, line 19 of the specification.

Please delete the paragraph beginning at page 10, line 20 of the specification.

Please delete the paragraph beginning at page 10, line 21 of the specification.

Please delete the paragraph beginning at page 10, line 22 of the specification.

Please delete the paragraph beginning at page 10, line 23 of the specification.

Please delete the paragraph beginning at page 10, line 24 of the specification.

Please delete the paragraph beginning at page 10, line 25 of the specification.

Please amend the paragraph beginning at page 16, line 11 of the specification to read as follows:

--Additional nucleic acid targets may be determined independently or can be selected from publicly available prokaryotic and eukaryotic genetic databases known to those skilled in the art. Preferred databases include, for example, Online Mendelian Inheritance in Man (OMIM), the Cancer Genome Anatomy Project (CGAP), GenBank, EMBL, PIR, SWISS-PROT, and the like. OMIM, which is a database of genetic mutations associated with disease, was developed, in part, for the National Center for Biotechnology Information (NCBI). OMIM is publicly available through the Internet at the world wide web at, for example, [ncbi.nlm.nih.gov/Omim/](http://ncbi.nlm.nih.gov/Omim/). CGAP, which is an interdisciplinary program to establish the information and technological tools required to decipher the molecular anatomy of a cancer cell. CGAP is publicly available through the Internet at the world wide web at, for example, [ncbi.nlm.nih.gov/ncicgap/](http://ncbi.nlm.nih.gov/ncicgap/). Some of these databases may contain complete or partial nucleotide sequences. In addition, nucleic acid targets can also be selected from private genetic databases. Alternatively, nucleic acid targets can be selected from available publications or

can be determined especially for use in connection with the present invention.--

Please amend the paragraph beginning at page 16, line 25 of the specification to read as follows:

--After a nucleic acid target is selected or provided, the nucleotide sequence of the nucleic acid target is determined and then compared to the nucleotide sequences of a plurality of nucleic acids from different taxonomic species. In one embodiment of the invention, the nucleotide sequence of the nucleic acid target is determined by scanning at least one genetic database or is identified in available publications. Preferred databases known and available to those skilled in the art include, for example, the Expressed Gene Anatomy Database (EGAD) and Unigene-Homo Sapiens database (Unigene), GenBank, and the like. EGAD contains a non-redundant set of human transcript (HT) sequences and is publicly available through the Internet at the world wide web at, for example, [tigr.org/tdb/egad/egad.html](http://tigr.org/tdb/egad/egad.html). Unigene is a system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters. Each Unigene cluster contains sequences that represent a unique gene, as well as related information such as the tissue types in which the gene has been expressed and map location.--

Please amend the paragraph beginning at page 17, line 10 of the specification to read as follows:

--In addition, Unigene contains hundreds of thousands of novel expressed sequence tag (EST) sequences. Unigene is publicly available through the Internet at the world wide web at, for example, [ncbi.nlm.nih.gov/UniGene/](http://ncbi.nlm.nih.gov/UniGene/). These databases can be used in connection with searching programs such as, for example, Entrez, which is known and available to those skilled in the art, and the like. Entrez is publicly available through the Internet at the world wide web at, for example, [ncbi.nlm.nih.gov/Entrez](http://ncbi.nlm.nih.gov/Entrez). Preferably, the most complete nucleic acid sequence representation available from various databases is used. The GenBank database, which is known and available to those skilled in the art, can also be used to obtain the most complete nucleotide sequence. GenBank is the NIH genetic sequence database and is an annotated collection of all publicly available DNA



sequences. GenBank is described in, for example, Nuc. Acids Res., 1998, 26, 1-7, which is incorporated herein by reference in its entirety, and can be accessed by those skilled in the art through the Internet at the world wide web at, for example, [ncbi.nlm.nih.gov/Web/Genbank/index.html](http://ncbi.nlm.nih.gov/Web/Genbank/index.html). Alternatively, partial nucleotide sequences of nucleic acid targets can be used when a complete nucleotide sequence is not available.--

Please amend the paragraph beginning at page 17, line 25 of the specification to read as follows:

--In another embodiment of the present invention, the nucleotide sequence of the nucleic acid target is determined by assembling a plurality of overlapping expressed sequence tags (ESTs). The EST database (dbEST), which is known and available to those skilled in the art, comprises approximately one million different human mRNA sequences comprising from about 500 to 1000 nucleotides, and various numbers of ESTs from a number of different organisms. dbEST is publicly available through the Internet at the world wide web at, for example, [ncbi.nlm.nih.gov/dbEST/index.html](http://ncbi.nlm.nih.gov/dbEST/index.html). These sequences are derived from a cloning strategy that uses cDNA expression clones for genome sequencing. ESTs have applications in the discovery of new genes, mapping of genomes, and identification of coding regions in genomic sequences. Another important feature of EST sequence information that is becoming rapidly available is tissue-specific gene expression data. This can be extremely useful in targeting selective gene(s) for therapeutic intervention. Since EST sequences are relatively short, they must be assembled in order to provide a complete sequence. Because every available clone is sequenced, it results in a number of overlapping regions being reported in the database.--

Please amend the paragraph beginning at page 18, line 9 of the specification to read as follows:

--Assembly of overlapping ESTs extended along both the 5' and 3' directions results in a full-length "virtual transcript." The resultant virtual transcript may represent an already characterized nucleic acid or may be a novel nucleic acid with no known biological function. The

Institute for Genomic Research (TIGR) Human Genome Index (HGI) database, which is known and available to those skilled in the art, contains a list of human transcripts. TIGR is publicly available through the Internet at the world wide web at, for example, [tigr.org/](http://tigr.org/). The transcripts were generated in this manner using TIGR-Assembler, an engine to build virtual transcripts and which is known and available to those skilled in the art. TIGR-Assembler is a tool for assembling large sets of overlapping sequence data such as ESTs, BACs, or small genomes, and can be used to assemble eukaryotic or prokaryotic sequences. TIGR-Assembler is described in, for example, Sutton, *et al.*, *Genome Science & Tech.*, **1995**, *1*, 9-19, which is incorporated herein by reference in its entirety, and is publicly available through the Internet via file transfer program at, for example [tigr.org/pub/software/TIGRAssembler](http://tigr.org/pub/software/TIGRAssembler). In addition, GLAXO-MRC, which is known and available to those skilled in the art, is another protocol for constructing virtual transcripts. In addition, "Find Neighbors and Assemble EST Blast" protocol, which runs on a UNIX platform, has been developed by Applicants to construct virtual transcripts. Preferred steps in the Find Neighbors and Assemble EST Blast protocol is described in the flowchart set forth in Figure 2. PHRAP is used for sequence assembly within Find Neighbors and Assemble EST Blast. PHRAP is publicly available through the Internet at, for example, [chimera.biotech.washington.edu/uwgc/tools/phrap.htm](http://chimera.biotech.washington.edu/uwgc/tools/phrap.htm). One skilled in the art can construct source code to carry out the preferred steps set forth in Figure 2.--

Please amend the paragraph beginning at page 19, line 23 of the specification to read as follows:

--Sequence similarity searches can be performed manually or by using several available computer programs known to those skilled in the art. Preferably, Blast and Smith-Waterman algorithms, which are available and known to those skilled in the art, and the like can be used. Blast is NCBI's sequence similarity search tool designed to support analysis of nucleotide and protein sequence databases. Blast is publicly available through the Internet at the world wide web at, for example, [ncbi.nlm.nih.gov/BLAST/](http://ncbi.nlm.nih.gov/BLAST/). The GCG Package provides a local version of Blast that can be used either with public domain databases or with any locally available searchable database. GCG Package v.9.0 is a commercially available software package that contains over 100 interrelated

software programs that enables analysis of sequences by editing, mapping, comparing and aligning them. Other programs included in the GCG Package include, for example, programs which facilitate RNA secondary structure predictions, nucleic acid fragment assembly, and evolutionary analysis. In addition, the most prominent genetic databases (GenBank, EMBL, PIR, and SWISS-PROT) are distributed along with the GCG Package and are fully accessible with the database searching and manipulation programs. GCG is publicly available through the Internet at the world wide web at, for example, [gcg.com/](http://gcg.com/). Fetch is a tool available in GCG that can get annotated GenBank records based on accession numbers and is similar to Entrez. Another sequence similarity search can be performed with GeneWorld and GeneThesaurus from Pangea. GeneWorld 2.5 is an automated, flexible, high-throughput application for analysis of polynucleotide and protein sequences. GeneWorld allows for automatic analysis and annotations of sequences. Like GCG, GeneWorld incorporates several tools for homology searching, gene finding, multiple sequence alignment, secondary structure prediction, and motif identification. GeneThesaurus 1.0tm is a sequence and annotation data subscription service providing information from multiple sources, providing a relational data model for public and local data.--

Please amend the paragraph beginning at page 20, line 24 of the specification to read as follows:

--Another toolkit capable of doing sequence similarity searching and data manipulation is SEALS, also from NCBI. This tool set is written in perl and C and can run on any computer platform that supports these languages. It is publicly available through the Internet at the world wide web at, for example, [ncbi.nlm.nih.gov/Walker/SEALS/](http://ncbi.nlm.nih.gov/Walker/SEALS/). This toolkit provides access to Blast2 or gapped blast. It also includes a tool called tax\_collector which, in conjunction with a tool called tax\_break, parses the output of Blast2 and returns the identifier of the sequence most homologous to the query sequence for each species present. Another useful tool is feature2fasta which extracts sequence fragments from an input sequence based on the annotation. An exemplary use for this tool is to create sequence files containing the 5' untranslated region of a cDNA sequence.--

Please amend the paragraph beginning at page 21, line 30 of the specification to read as follows:

--In another embodiment of the invention, the sequences required are obtained by searching ortholog databases. One such database is Hovergen, which is a curated database of vertebrate orthologs. Ortholog sets may be exported from this database and used as is, or used as seeds for further sequence similarity searches as described above. Further searches may be desired, for example, to find invertebrate orthologs. Hovergen is publicly available through the Internet via file transfer program at, for example, [pbil.univ-lyon1.fr/pub/hovergen/](http://pbil.univ-lyon1.fr/pub/hovergen/). A database of prokaryotic orthologs, COGS, is available and can be used interactively through the Internet at the world wide web at, for example, [ncbi.nlm.nih.gov/COG/](http://ncbi.nlm.nih.gov/COG/).--

Please amend the paragraph beginning at page 23, line 22 of the specification to read as follows:

--Sequence homology between the window sequence of the target nucleic acid and the query sequence of any of the plurality of nucleic acid sequences obtained as described above, is preferably at least 60%, more preferably at least 70%, more preferably at least 80%, and most preferably at least 90%. The most preferable method of choosing the threshold is to have the computer automatically try all thresholds from 50% to 100% and choose a threshold based a metric provided by the user. One such metric is to pick the threshold such that exactly n hits are returned, where n is usually set to 3. This process is repeated until every base on the query nucleic acid, which is a member of the plurality of nucleic acids described above, has been compared to every base on the master target sequence. The resulting scoring matrix can be plotted as a scatter plot. Based on the match density at a given location, there may be no dots, isolated dots, or a set of dots so close together that they appear as a line. The presence of lines, however small, indicates primary sequence homology. Sequence conservation within nucleic acid molecules, particularly the UTRs of RNA, in divergent species is likely to be an indicator of conserved regulatory elements that are also likely to have a secondary structure. The results of

the interspecies sequence comparison can be analyzed using MS Excel and visual basic tools in an entirely automated manner as known to those skilled in the art.--

Please amend the paragraph beginning at page 23, line 22 of the specification to read as follows:

--In another embodiment of the invention, secondary structure analysis is performed by self complementarity comparison. Preferably, self complementarity comparison is performed using Compare, described above. More preferably, Compare can be modified to expand the pairing matrix to account for G-U or U-G basepairs in addition to the conventional Watson-Crick G-C/C-G or A-U/U-A pairs. Such a modified Compare program (modified Compare) begins by predicting all possible base-pairings within a given sequence. As described above, a small but conserved region, preferably a UTR, is identified based on primary sequence comparison of a series of orthologs. In modified Compare, each of these sequences is compared to its own reverse complement. Allowable base-pairings include Watson-Crick A-U, G-C pairing and non-canonical G-U pairing. An overlay of such self complementarity plots of all available orthologs, and selection for the most repetitive pattern in each, results in a minimal number of possible folded configurations. These overlays can then used in conjunction with additional constraints, including those imposed by energy considerations described above, to deduce the most likely secondary structure.--

Please amend the paragraph beginning at page 24, line 17 of the specification to read as follows:

--In one embodiment of the invention, secondary structure analysis is performed by alignment and covariance analysis. Numerous protocols for alignment and covariance analysis are known to those skilled in the art. Preferably, alignment is performed by ClustalW, which is available and known to those skilled in the art. ClustalW is a tool for multiple sequence alignment that, although not a part of GCG, can be added as an extension of the existing GCG tool set and used with local sequences. ClustalW is publicly available through the Internet at, for example,

dot.imgen.bcm.tmc.edu:9331/multialign/Options/clustalw.html. ClustalW is also described in Thompson, *et al.*, *Nuc. Acids Res.*, **1994**, 22, 4673-4680, which is incorporated herein by reference in its entirety. These processes can be scripted to automatically use conserved UTR regions identified in earlier steps. Seqed, a UNIX command line interface available and known to those skilled in the art, allows extraction of selected local regions from a larger sequence. Multiple sequences from many different species can be clustered and aligned for further analysis.--

Please amend the paragraph beginning at page 25, line 11 of the specification to read as follows:

--Covariation is a process of using phylogenetic analysis of primary sequence information for consensus secondary structure prediction. Covariation is described in the following references, each of which is incorporated herein by reference in their entirety: Gutell, *et al.*, "Comparative Sequence Analysis Of Experiments Performed During Evolution" In Ribosomal RNA Group I Introns, Green, Ed., Austin:Landes, **1996**; Gautheret, *et al.*, *Nuc. Acids Res.*, **1997**, 25, 1559-1564; Gautheret, *et al.*, *RNA*, **1995**, 1, 807-814; Lodmell, *et al.*, *Proc. Natl. Acad. Sci. USA*, **1995**, 92, 10555-10559; Gautheret, *et al.*, *J. Mol. Biol.*, **1995**, 248, 27-43; Gutell, *Nuc. Acids Res.*, **1994**, 22, 3502-3517; Gutell, *Nuc. Acids Res.*, **1993**, 21, 3055-3074; Gutell, *Nuc. Acids Res.*, **1993**, 21, 3051-3054; Woese, *Proc. Natl. Acad. Sci. USA*, **1989**, 86, 3119-3122; and Woese, *et al.*, *Nuc. Acids Res.*, **1980**, 8, 2275-2293. Preferably, covariance software is used for covariance analysis. Preferably, Covariation, a set of programs for the comparative analysis of RNA structure from sequence alignments, is used. Covariation uses phylogenetic analysis of primary sequence information for consensus secondary structure prediction. Covariation is publicly available through the Internet at the world wide web at, for example [mbio.ncsu.edu/RNaseP/info/programs/programs.html](http://mbio.ncsu.edu/RNaseP/info/programs/programs.html). A complete description of a version of the program has been published (Brown, J. W. 1991 Phylogenetic analysis of RNA structure on the Macintosh computer. CABIOS7:391-393). The current version is v4.1, which can perform various types of covariation analysis from RNA sequence alignments, including standard covariation analysis, the identification of compensatory base-changes, and mutual information analysis. The program is well-documented and comes with extensive

example files. It is compiled as a stand-alone program; it does not require Hypercard (although a much smaller 'stack' version is included). This program will run in any Macintosh environment running MacOS v7.1 or higher. Faster processor machines (68040 or PowerPC) is suggested for mutual information analysis or the analysis of large sequence alignments.--

Please amend the paragraph beginning at page 26, line 5 of the specification to read as follows:

--In another embodiment of the invention, secondary structure analysis is performed by secondary structure prediction. There are a number of algorithms that predict RNA secondary structures based on thermodynamic parameters and energy calculations. Preferably, secondary structure prediction is performed using either M-fold or RNA Structure 2.52. M-fold is publicly available through the Internet at the world wide web at, for example, [ibc.wustl.edu/~zucker/ma/form2.cgi](http://ibc.wustl.edu/~zucker/ma/form2.cgi) or can be downloaded for local use on UNIX platforms. M-fold is also available as a part of GCG package. RNA Structure 2.52 is a windows adaptation of the M-fold algorithm and is publicly available through the Internet at, for example, [128.151.176.70/RNAstructure.html](http://128.151.176.70/RNAstructure.html).--

Please amend the paragraph beginning at page 26, line 29 of the specification to read as follows:

--In another preferred embodiment of the invention, the output of AlignHits is read by a program called RevComp. A block diagram of this program is shown in Figure 9. This program could be reproduced by one skilled in the art. A preferred purpose of this program is to use base pairing rules and ortholog evolution to predict RNA secondary structure. RNA secondary structures are composed of single stranded regions and base paired regions, called stems. Since structure conserved by evolution is searched, the most probable stem for a given alignment of ortholog sequences is the one which could be formed by the most sequences. Possible stem formation or base pairing rules is determined by, for example, analyzing base pairing statistics of stems which have been determined by other techniques such as NMR. The output of RevComp is a sorted list of possible structures, ranked by the percentage of ortholog set member sequences

which could form this structure. Because this approach uses a percentage threshold approach, it is insensitive to noise sequences. Noise sequences are those that either not true orthologs, or sequences that made it into the output of AlignHits due to high sequence homology even though they do not represent an example of the structure which is searched. A very similar algorithm is implemented using Visual basic for Applications (VBA) and Microsoft Excel to be run on PCs, to generate the reverse complement matrix view for the given set of sequences.--

Please amend the paragraph beginning at page 28, line 13 of the specification to read as follows:

--After the hypothetical structure motifs are determined from the secondary structure analysis described above, a family of structure descriptor elements is constructed. Preferably, the structural motifs described above are converted into a family of descriptor elements. An exemplary descriptor element is shown in Figure 6. One skilled in the art is familiar with construction of descriptors. Structure descriptors are described in, for example, Laferriere, et al., Comput. Appl. Biosci., 1994, 10, 211-212, incorporated herein by reference in its entirety. A different structure descriptor element is constructed for each of the structural motifs identified from the secondary structure analysis. Briefly, the secondary structure is converted to a generic text string, such as shown in Figure 6. For novel motifs, further biochemical analysis such as chemical mapping or mutagenesis may be needed to confirm structure predictions. Descriptor elements may be defined to have various stringency.--

Please amend the paragraph beginning at page 28, line 24 of the specification to read as follows:

--For example, referring to Figure 6, the region termed H1, which comprises the first region of the stem, can be described as NNN:NNN, which contemplates any complementary base pairing including G-C, C-G, A-U, and U-A. The H1 region may also be designated so as to include only C-G or A-U, etc., base pairing. In addition, the descriptor elements can be defined to allow for a wobble. Thus, descriptor elements can be defined to have any level of stringency



desired by the user. Applicants' invention, thus, is also directed to a database comprising different descriptor elements. --

Please amend the paragraph beginning at page 29, line 10 of the specification to read as follows:

--In one embodiment of the invention, nucleic acids having secondary structure which correspond to the structure descriptor elements are identified by searching at least one database. Any genetic database can be searched. Preferably, the database is a UTR database, which is a compilation of the untranslated regions in messenger RNAs. A UTR database is publicly available through the Internet via file transfer program at [area.ba.cnr.it/pub/embnet/database/utr/](http://area.ba.cnr.it/pub/embnet/database/utr/). Preferably the database is searched using a computer program, such as, for example, Rnamot, a UNIX-based motif searching tool available from Daniel Gautheret. Each "new" sequence that has the same motif is then queried against public domain databases to identify additional sequences. Results are analyzed for recurrence of pattern in UTRs of these additional ortholog sequences, as described below, and a database of RNA secondary structures is built. One skilled in the art is familiar with Rnamot. Briefly, Rnamot takes a descriptor string, such as the one shown in Figure 6, and searches any Fasta format database for possible matches. Descriptors can be very specific, to match exact nucleotide(s), or can have built-in degeneracy. Lengths of the stem and loop can also be specified. Single stranded loop regions can have a variable length. G-U pairings are allowed and can be specified as a wobble parameter. Allowable mismatches can also be included in the descriptor definition. Functional significance is assigned to the motifs if their biological role is known based on previous analysis. Known regulatory regions such as Iron Response Element have been found using this technique (see, Example 1 below). In embodiments of the invention in which a database containing prokaryotic molecular interaction sites is compiled, it is preferable to refrain from searching human sequences or, alternatively, discarding human sequences when found.--

**DOCKET NO.: IBIS-0012**

Please amend the paragraph beginning at page 32, line 26 of the specification to read as follows:

--An early step in the process is to use the master sequence (nucleotide or protein) to find and rank related sequences in the database (orthologs and paralogs). Sequence similarity search algorithms are used for this purpose. All sequence similarity algorithms calculate a quantitative measure of similarity for each result compared with the master sequence. An example of a quantitative result is an E-value obtained from the Blast algorithm. The E-values for a blast search of the non-redundant GenBank database using ferritin mRNA as the query sequence illustrates the use of quantitative analysis of sequence similarity searches. The E-value is the probability that a match between a query sequence and a database sequence occurs due to random chance. Therefore, the lower an E-value the more likely that two sequences are truly related. A plot of the lowest E-value scores for ferritin is shown in Figure 7. Sequences that meet the cutoff criteria are selected for more detailed comparisons according to a set of rules described below. Since an objective of the sequence similarity search to find distantly related orthologs and paralogs, it is preferable that the cutoff criteria not be too stringent, or the target of the search may be excluded.--

Please amend the paragraph beginning at page 33, line 23 of the specification to read as follows:

--When the human mRNA and mouse mRNA sequences for ferritin, which each contain an IRE in the 5'-UTR, are analyzed in this manner, a plot showing the regions of sequence similarity is produced. Pairwise analysis of the human and mouse ferritin mRNA sequences illustrate several important aspects of this type of analysis. Regions of each mRNA that encode the amino acid sequence have the highest degree of similarity, while the untranslated regions are less similar. In both the human and mouse ferritin mRNAs the IREs are located in the extreme 5' end of each mRNA. This demonstrates an important point -- the sequence conservation in the region of the IRE structure does not stand out against the background of sequence similarity between the human and mouse ferritin sequences. In contrast, in the comparison of human and

trout or human and chicken ferritin mRNAs, the IREs can be immediately identified. This is because the sequence of the UTRs between human and trout or human and chicken are separated by greater evolutionary distance than human and mouse, which is logical in view of the evolutionary distance that separates humans from birds and fish compared with other mammals. Comparing the human sequence to that of birds and fish is informative because the natural drift due to evolution has allowed many sequence changes in the UTRs. However, the IRE sequences are more constrained because they form an important structure. Thus, they stand out better and can be more readily identified.--

Please amend the paragraph beginning at page 34, line 28 of the specification to read as follows:

--The software used in the present invention makes the decision whether or not to compare sequences pairwise using a lookup table based upon the evolutionary distances between species: An example of a small lookup table using the examples described above is shown in Figure 8. The lookup table in the present invention includes all species that have sequences deposited in GenBank. Q-Compare in conjunction with CompareOverWins decides which sequences to compare pairwise.--

Please amend the paragraph beginning at page 35, line 4 of the specification to read as follows:

--Sets of sequences that show evidence of conservation in orthologs and paralogs or other related genes are analyzed for the ability to form internal structure. This is accomplished by analyzing each sequence in a matrix where the sequence is plotted 5' to 3' on the X axis and its reverse complement is plotted 5' to 3' on the Y axis, such as in, for example, self-complementary analysis. Matches that correspond to potential intramolecular base pairs are scored according to a table of values. When the human ferritin IRE sequence is analyzed in this fashion, the diagonals indicate potential self-complementary regions. Each of the 13 IRE sequences described in this example were analyzed in the same fashion. While each of the sequences can form a variety of

different structures, the structure most likely to occur is one common to all the sequences. By superimposing the plots of all 13 individual sequences, the potential structure common to all the sequences is deduced.--

Please amend the paragraph beginning at page 36, line 5 of the specification to read as follows:

--Hovergen was used to identify related sequences at the species and order levels. Sequences corresponding to each of these orthologs was saved in GenBank format and grouped together in a single data file. Untranslated regions in both the 5' and 3' flanks of the coding region was extracted using SEALS and COWX, as shown in Figure 11.--

Please amend the paragraph beginning at page 36, line 11 of the specification to read as follows:

--The IRE sequences are more constrained because they form an important structure. Thus, they stand out better and can be more readily identified even in closely related sequences. However, for this to work for any gene, the compare algorithm has been rewritten (see, Figures 5A-C). This new tool, CompareOverWins, allows a dynamic selection of both the range of window sizes, as well the hit threshold. This algorithm needs as its input parsed and separated 5' and 3' UTR sequences. We use tools available within the Seals genome analysis package described earlier to achieve this. Figure 11 describes the steps involved.--

Please amend the paragraph beginning at page 36, line 19 of the specification to read as follows:

--To identify the IRE using the methods described herein, the compare over windows algorithm was used and the results visualized using AlignHits (Figure 5D for the algorithm). In addition to optimizing the thresholding, CompareOverWins also extracts the sequence corresponding to the hits. ClustalW (version 1.74) was used on the extracted sequences to create

a locally gapped alignment. A representative flow scheme for this approach is shown in Figure 13.--

Please amend the paragraph beginning at page 36, line 27 of the specification to read as follows:

--Sets of sequences that show evidence of conservation in orthologs and paralogs or other related genes are analyzed for the ability to form internal structure. This is accomplished by analyzing each sequence in a matrix where the sequence is plotted 5' to 3' on the X axis and its complement is plotted 5' to 3' on the Y axis, such as in, for example, self-complementary analysis. Matches that correspond to potential intramolecular base pairs are scored according to a table of values. When the human ferritin IRE sequence is analyzed in this fashion, the diagonals indicate potential self-complementary regions. Each of the 13 IRE sequences described in this example were analyzed in the same fashion. While each of the sequences can form a variety of different structures, the structure most likely to occur is one common to all the sequences. By superimposing the plots of all 13 individual sequences, the potential structure common to all the sequences is deduced.--

Please amend the paragraph beginning at page 37, line 9 of the specification to read as follows:

--The above scheme has been implemented algorithmically into a program called RevComp (see, Figure 9). RevComp creates a sorted list of all the structures. Representative results can be viewed either as a "dome" output or as a "connect" or "ct" file which can be used in one of many RNA structure viewing programs (RNAstructure, RNAViz, etc.). A representative example of such a structure drawing is shown in Figure 14.--

Please amend the paragraph beginning at page 37, line 21 of the specification to read as follows:

--Phylogenetic tree outputs for all Histone orthologs in Hovergen database was obtained. Each of these orthologs was saved in GenBank format and grouped together in a single data file. Untranslated regions in both the 5' and 3' flanks of the coding regions were extracted and compared using SEALS and COWX as described earlier (*see*, Figures 11 and 13). --

Please amend the paragraph beginning at page 37, line 26 of the specification to read as follows:

--Following extraction and comparison by SEALS and COWX, Align Hits was used to determine potentially interesting regions. One such region is shown encircled]. The sequences corresponding to the region of interest was extracted from all species for alignment with CLUSTAL W (1.74). Following extraction of sequence information from Align Hits, CLUSTAL W (1.74) was used to provide multiple sequence alignment shown. Each of the putative hit sequences was analyzed for the ability to form internal structure. This was accomplished by analyzing each sequence in a matrix where the sequence was plotted 5' to 3' on the X axis and its complement is plotted 5' to 3' on the Y axis. Base-pairs along the diagonals indicate potential self-complementary regions that can form secondary structures. A representative sequence alignment in a dome format can show potential stem formation between the base pairs. Following conversion of the dome format file to a ct file, RNA Structure 3.21 is used to visualize the structure.--

Please amend the paragraph beginning at page 38, line 10 of the specification to read as follows:

--Vimentin is an intermediate filament protein whose 3'UTR is highly conserved between species. Previous studies by Zehner et al., (Nuc. Acids Res., 1997, 25, 3362-3370) has shown that a proposed a complex stem-loop structure contained within this region may be important for vimentin mRNA functions such as mRNA localization. The same region was identified using the present analysis, thus validating the present approach. In addition, based on the analyses described herein, a second stem-loop structure that occurs downstream of the previously

DOCKET NO.: IBIS-0012

proposed structure that may have a role in regulating vimentin fuction as well has been identified.--

Please amend the paragraph beginning at page 38, line 18 of the specification to read as follows:

--A representative phylogenetic tree output for all Vimentin orthologs in Hovergen database was obtained. Each of these orthologs was saved in GenBank format and grouped together in a single data file. Untranslated regions in both the 5' and 3' flanks of the coding regions were extracted and compared using SEALS and COWX as described earlier (*see*, Figures 11 and 13). --

Please amend the paragraph beginning at page 38, line 23 of the specification to read as follows:

--Following extraction and comparison by SEALS and COWX, Align Hits was used to determine potentially interesting regions. Two such regions appeared, and were used for subsequent analyses. Following extraction of sequence information from Align Hits for the first region, CLUSTAL W was used to provide multiple sequence alignment. Potential stem formation between base pairs was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure. This structure is very similar to the one proposed by Zehner *et al.*. Zehner *et al.* presented a detailed chemical analysis of their proposed structure for the minimal binding domain in the 3' UTR of Vimentin. This analysis included cleavage with single-strand-specific (ChS or T1) or double-strand-specific (V1) nucleases as well as after exposure to lead acetate.--

Please amend the paragraph beginning at page 39, line 5 of the specification to read as follows:

--Following extraction of sequence information from Align Hits for the second region, CLUSTAL W was used to provide multiple sequence alignment. The potential stem formation between base pairs in the second region was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure for the second region.--

Please amend the paragraph beginning at page 39, line 18 of the specification to read as follows:

--A representative phylogenetic tree output for all Transferrin receptor orthologs in Hovergen database was obtained. Each of these orthologs was saved in GenBank format and grouped together in a single data file. Untranslated regions in both the 5' and 3' flanks of the coding region were extracted and compared using SEALS and COWX as described earlier (*see*, Figures 11 and 13).--

Please amend the paragraph beginning at page 39, line 23 of the specification to read as follows:

--Following extraction and comparison by SEALS and COWX, Align Hits was used to determine potentially interesting regions. The first region, between base pairs 920 to 990, in the 3 prime UTR of transferrin receptor was extracted from all species for alignment with CLUSTAL W (1.74).--

Please amend the paragraph beginning at page 39, line 28 of the specification to read as follows:

--Following extraction of sequence information from Align Hits for the first region, CLUSTAL W (1.74) was used to provide multiple sequence alignment. A representative potential stem formation between base pairs was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure. This can be seen where a vertical line intersects a series of horizontal



lines representing sequence information from a set of species. This] The second region, between base pairs 990 to 1050, in the 3 prime UTR of transferrin receptor was extracted from all species for alignment with CLUSTAL W (1.74).--

Please amend the paragraph beginning at page 40, line 7 of the specification to read as follows:

--Following extraction of sequence information from Align Hits for the second region, CLUSTAL W (1.74) was used to provide multiple sequence alignment. Potential stem formation between base pairs was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure. Following extraction and comparison by SEALS and COWX, Align Hits was used to determine potentially interesting regions. The third region, between base pairs 1372 to 1423, in the 3 prime UTR of transferrin receptor was extracted from all species for alignment with CLUSTAL W (1.74).--

Please amend the paragraph beginning at page 40, line 17 of the specification to read as follows:

--Following extraction of sequence information from Align Hits for the third region, CLUSTAL W (1.Ex.34) was used to provide multiple sequence alignment. Potential stem formation between base pairs was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure. Following extraction and comparison by SEALS and COWX, Align Hits was used to determine potentially interesting regions. The fourth region, between base pairs 1439 to 1479, in the 3 prime UTR of transferrin receptor was extracted from all species for alignment with CLUSTAL W (1.74).--

Please amend the paragraph beginning at page 40, line 27 of the specification to read as follows:

--Following extraction of sequence information from Align Hits for the fourth region, CLUSTAL W (1.Ex.34) was used to provide multiple sequence alignment. Potential stem formation between base pairs was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure. Following extraction and comparison by SEALS and COWX, Align Hits was used to determine potentially interesting regions. The fifth region, between base pairs 1479 to 1542, in the 3 prime UTR of transferrin receptor was extracted from all species for alignment with CLUSTAL W (1.74).--

Please amend the paragraph beginning at page 41, line 6 of the specification to read as follows:

--Following extraction of sequence information from Align Hits for the fifth region, CLUSTAL W (1.Ex.34) was used to provide multiple sequence alignment. Potential stem formation between base pairs was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure.--

Please amend the paragraph beginning at page 41, line 12 of the specification to read as follows:

--Ornithine decarboxylase (ODC) is the first enzyme in the polyamine biosynthetic pathway. Studies have shown existence of translational regulatory elements both in the 5' and 3' untranslated regions (Grens et al., J. Biol. Chem., 1990, 265, 11810). Secondary structures have been proposed to exist in both these regions, though there is no conclusive evidence for it. The methods described herein identified two structures in the 3' UTR, as shown below. The presence of one of these structures (*see*, Figure 15) was verified using mass spectrometry probing (Griffey, et al., Proc. SPIE-Int. Soc. Opt. Eng., 2985 (Ultrasensitive Biochemical Diagnostics II): 82-86, which is incorporated herein by reference in its entirety). Two representative sequences that showed slight variation in their lengths were made into RNA and subjected to MS structure

probing. Results shown in Figure 15 confirm the presence of a stem-loop structure. Accordingly, identification of a novel secondary structure can be identified from the methods described herein, and such existence has been independently verified by structure probing.--

Please amend the paragraph beginning at page 41, line 25 of the specification to read as follows:

--Phylogenetic tree outputs for all Ornithine Decarboxylase orthologs in Hovergen database were obtained. Each of these orthologs was saved in GenBank format and grouped together in a single data file. Untranslated regions in both the 5' and 3' flanks of the coding region were extracted and compared using SEALS and COWX as described earlier (*see*, Figures 11 and 13).--

Please amend the paragraph beginning at page 42, line 1 of the specification to read as follows:

--Following extraction and comparison by SEALS and COWX, Align Hits was used to determine potentially interesting regions. Two such regions appeared, and were used for subsequent analyses. Following extraction of sequence information from the first region, CLUSTAL W (1.74) was used to provide multiple sequence alignment shown. Each of the putative hit sequences was analyzed for the ability to form internal structure in a reverse complement matrix. This was accomplished by analyzing each sequence in a matrix where the sequence is plotted 5' to 3' on the X axis and its complement is plotted 5' to 3' on the Y axis. Base-pairs along the diagonals indicate potential self- complementary regions that can form secondary structures. Domes view of the potential stem formation between base pairs in region 1 is given above the sequence alignment was determined using RevComp. RNA Structure 3.2 was used to visualize the structure.--

Please amend the paragraph beginning at page 42, line 13 of the specification to read as follows:

--Mass spectrometry analyses techniques were used to probe for structure. The cluster alignment of the first region of ornithine decarboxylase 3' UTR showed presence of gaps/inserts in the multiple alignment. Two representative RNAs (gi404561 and gi35135) from the alignments were used for this experiment. Analysis of the pattern of induced fragmentation showed a very strong likelihood for base-pairing along the top half of the stem-loop structure. This corresponds to bases 11-14 and 20-23 in 404561 or bases 8-11 and 18-21 in 35135. Bulged bases (G9 in 404561 or U22 in 35135) also showed characteristic fragmentation pattern. The bottom-half of the structure appeared to be less stable, and showed some fragmentation where our analyses had predicted base-pairing. This was particularly true in the sequence 35135. This region, however, has several contiguous A-U or G-U base-pairs which tend to be less stable, and therefore have a higher probability of fragmentation.--

Please amend the paragraph beginning at page 42, line 24 of the specification to read as follows:

--Following extraction of sequence information from Align Hits for the second region, CLUSTAL W was used to provide multiple sequence alignment. Potential stem formation between base pairs in the second region was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure for the second region.--

Please amend the paragraph beginning at page 43, line 2 of the specification to read as follows:

--A representative phylogenetic tree output for all IL-2 orthologs in Hovergen database was obtained. Each of these orthologs was saved in GenBank format and grouped together in a single data file. Untranslated regions in both the 5' and 3' flanks of the coding region were extracted and compared using SEALS and COWX as described earlier (*see*, Figures 11 and 13.--

Please amend the paragraph beginning at page 43, line 7 of the specification to read as follows:

--Following extraction and comparison by SEALS and COWX, Align Hits was used to determine potentially interesting regions in the 3'UTR region. Two such regions appear, and were used for subsequent analyses. Following extraction of sequence information from Align Hits for the first region, CLUSTAL W (1.74) was used to provide multiple sequence alignment. Domes view of the potential stem formation between base pairs in the first region was given above the sequence alignment using RevComp. RNA Structure 3.2 was used to visualize the structure. Following extraction of sequence information from Align Hits for the second region, CLUSTAL W (1.74) was used to provide multiple sequence alignment. Potential stem formation between base pairs in the second region was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure for the second region.--

Please amend the paragraph beginning at page 43, line 20 of the specification to read as follows:

--In addition to the two regions described above, a third region, downstream of, and partially overlapping the second region, was identified using an alternate reference sequence (3087784.fa). Following extraction of sequence information from Align Hits for this region, CLUSTAL W (1.74) was used to provide multiple sequence alignment. Potential stem formation between base pairs in the third region was shown above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure for the third region.--

Please amend the paragraph beginning at page 43, line 29 of the specification to read as follows:

--Representative phylogenetic tree output for all IL-4 orthologs in Hovergen database was obtained. Each of these orthologs was saved in GenBank format and grouped together in a single

data file. Untranslated regions in both the 5' and 3' flanks of the coding region were extracted and compared using SEALS and COWX as described earlier (*see*, Figures 11 and 13). --

Please amend the paragraph beginning at page 44, line 4 of the specification to read as follows:

--Following extraction and comparison by SEALS and COWX, Align Hits was used to determine potentially interesting regions in the 5'UTR region. Following extraction of sequence information from Align Hits for the above region, CLUSTAL W (1.74) was used to provide multiple sequence alignment. Domes view of the potential stem formation between base pairs in the region was given above the sequence alignment using RevComp. RNA Structure 3.2 was used to visualize the structure.--

Please amend the paragraph beginning at page 44, line 11 of the specification to read as follows:

--Align Hits was used to view hits in the 3'UTR region of IL-4. Following extraction of sequence information from Align Hits for the 3' UTR region, CLUSTAL W (1.74) was used to provide multiple sequence alignment. Potential stem formation between base pairs in the second region was given above the sequence alignment in a dome format. Following conversion of the dome format file to a ct file, RNA Structure 3.21 was used to visualize the structure for the second region.--

**In the Claims:**

Please amend claims 35, 43 and 51 to read as follows.

35. (Twice Amended) An oligonucleotide comprising a molecular interaction site that is present in the RNA of a selected organism and in the RNA of at least one additional organism, wherein said molecular interaction site serves as a binding site for at least one molecule that when bound to said molecular interaction site modulates the expression of said RNA in said selected organism, wherein